机器学习与量化交易实战

第六课

R^2

density Estimat

Intro statistical LEARNING

$\frac{\partial y}{\partial w}$

10

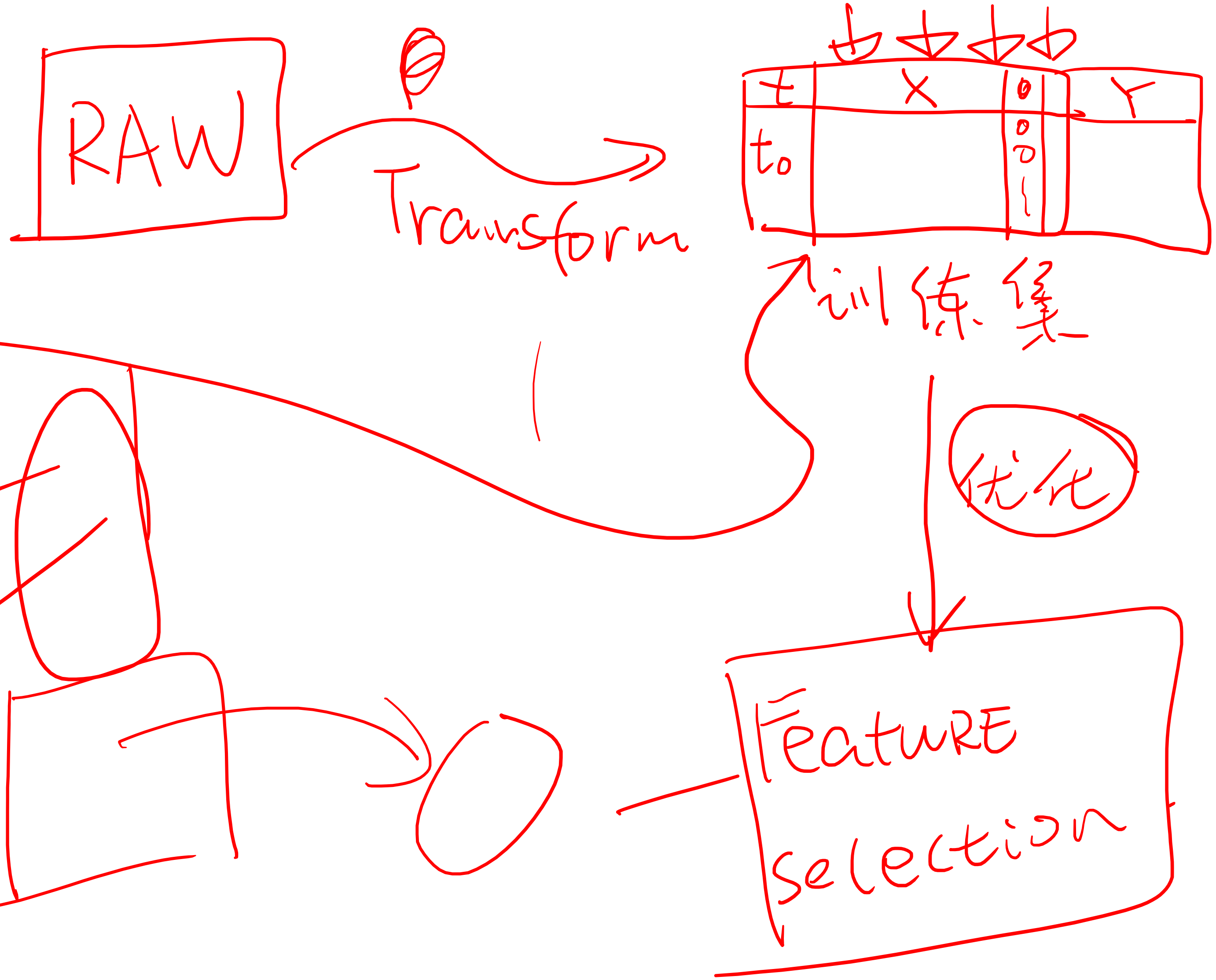
X10

Rmsprop

SGD

Outline

- 特征选择
- 遗传算法
- 深入理解BP算法
- RNN



RMSI

上次作业cont.

周六 最后一张PPT

最后一个数据集

benchmark R^2 on

KNN

And OLS

3/7 70% 30%

$\hat{Y}$ Y

Tx Y

Real Y

R^2

Regression

Try:

Feature Selection（today’s topic) + some other models

Sklearn

..and report the best performance combinations..

X [X_1 – αX_i — X_p] $P = 300$, 1000, 200

- *Subset Selection*. We identify a subset of the p predictors that we believe to be related to the response. We then fit a model using least squares on the reduced set of variables.

30

- *Shrinkage*. We fit a model involving all p predictors, but the estimated coefficients are shrunken towards zero relative to the least squares estimates. This shrinkage (also known as *regularization*) has the effect of reducing variance and can also perform variable selection.

LASSO

- *Dimension Reduction*. We project the p predictors into a M-dimensional subspace, where $M < p$. This is achieved by computing M different *linear combinations*, or *projections*, of the variables. Then these M projections are used as predictors to fit a linear regression model by least squares.

$f(X_{300\times 1}) \rightarrow Z_{20}$

Subset Selection

300

1 2 3

$50\binom{300}{2}$ $\binom{300}{3}$

5

Best subset and stepwise model selection procedures

Best Subset Selection

1. Let $\mathcal{M}_0$ denote the *null model*, which contains no predictors. This model simply predicts the sample mean for each observation.
2. For $k = 1, 2, \ldots p$:
   (a) Fit all $\binom{p}{k}$ models that contain exactly k predictors.
   (b) Pick the best among these $\binom{p}{k}$ models, and call it $\mathcal{M}_k$. Here *best* is defined as having the smallest RSS, or equivalently largest R^2.
3. Select a single best model from among $\mathcal{M}_0, \ldots, \mathcal{M}_p$ using cross-validated prediction error, C_p (AIC), BIC, or adjusted R^2.

Forward Stepwise Selection

- Forward stepwise selection begins with a model containing no predictors, and then adds predictors to the model, one-at-a-time, until all of the predictors are in the model.
- In particular, at each step the variable that gives the greatest *additional* improvement to the fit is added to the model.

$\{X_1 \cdots X_8 \; X_p\}_{300}$

In Detail

Backward

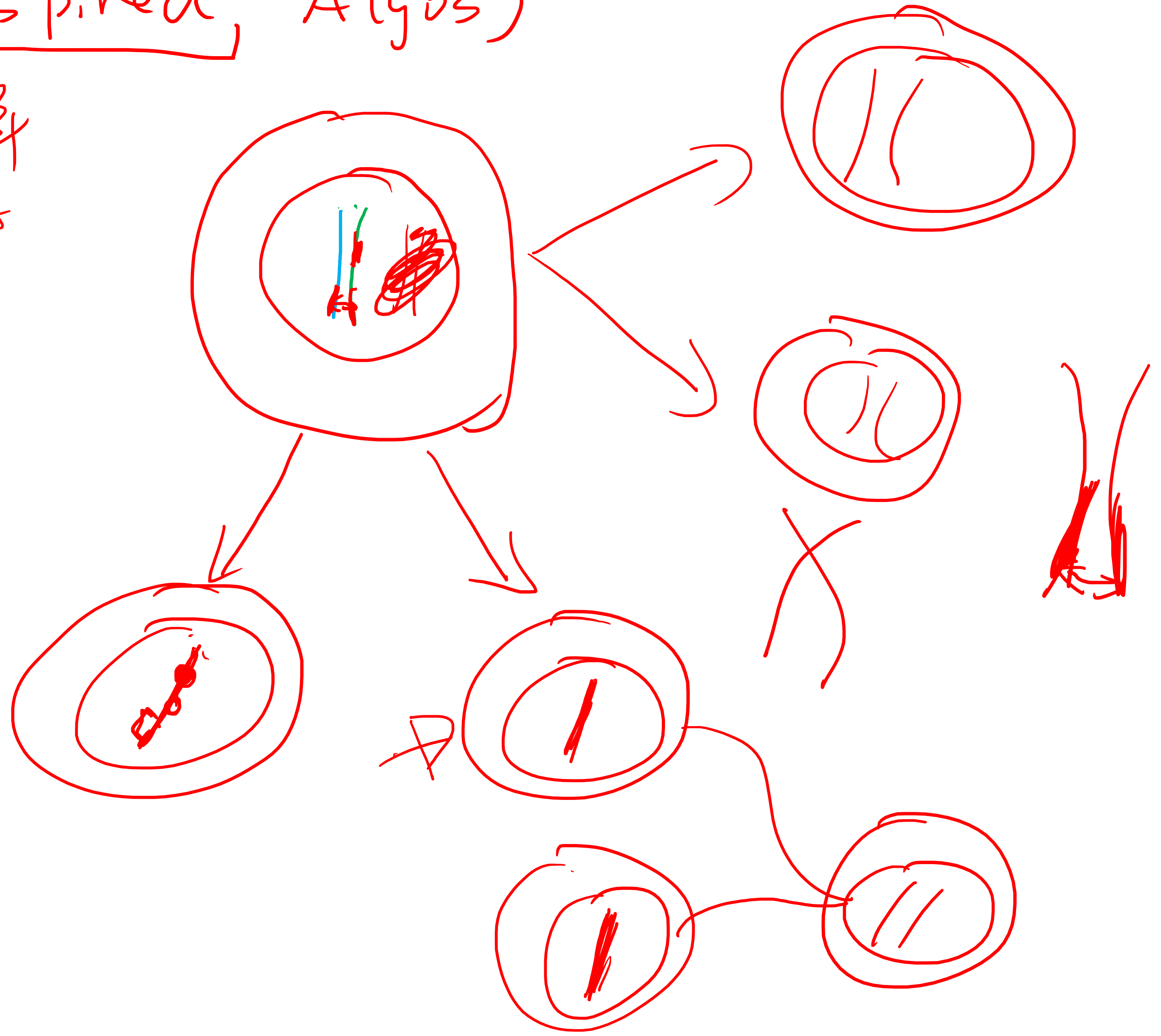
$\{X_8\}$

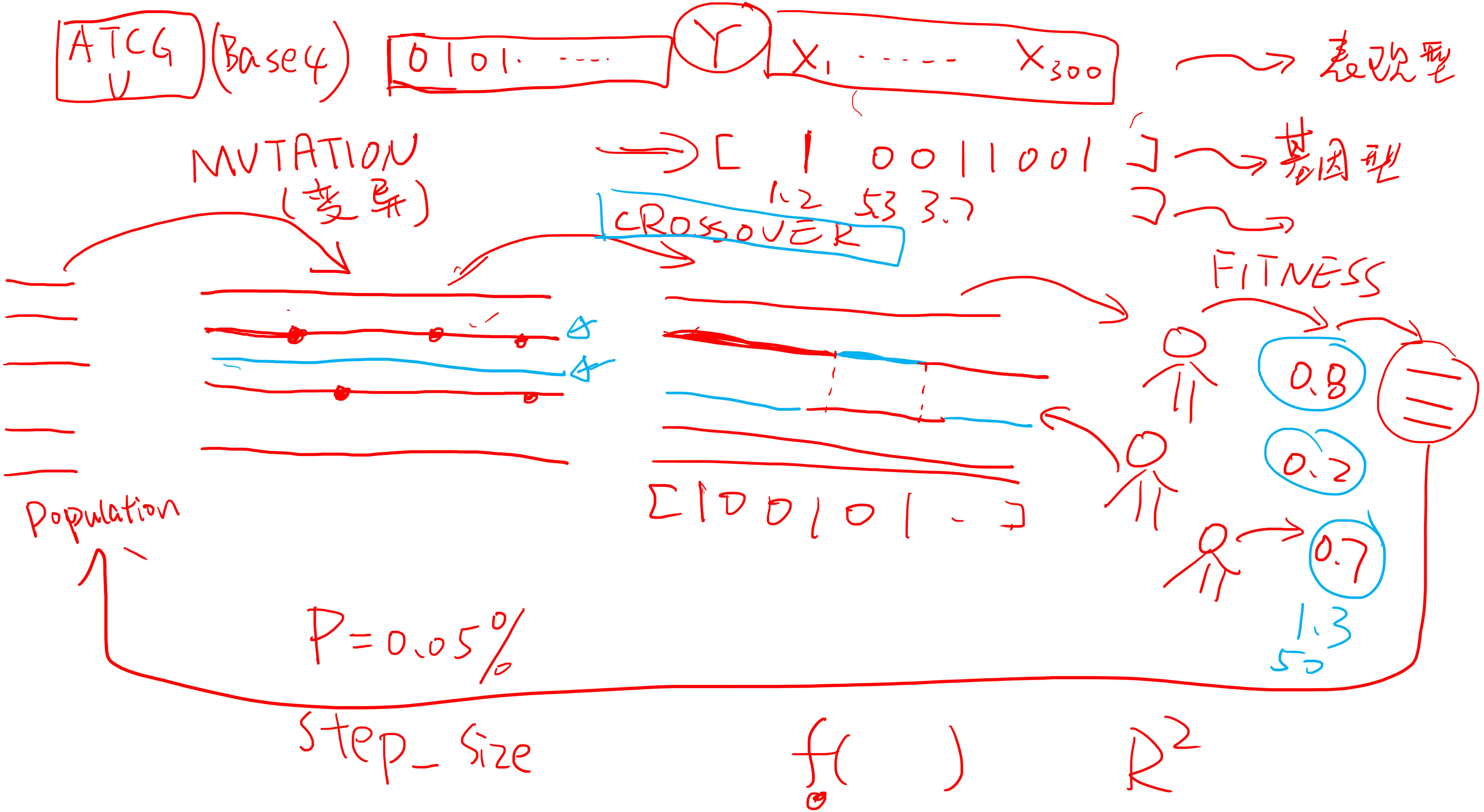
Forward Stepwise Selection

1. Let $\mathcal{M}_0$ denote the *null* model, which contains no predictors.
2. For $k = 0, \ldots, p - 1$:
   - 2.1 Consider all $p - k$ models that augment the predictors in $\mathcal{M}_k$ with one additional predictor.
   - 2.2 Choose the *best* among these $p - k$ models, and call it $\mathcal{M}_{k+1}$. Here *best* is defined as having smallest RSS or highest R^2.
3. Select a single best model from among $\mathcal{M}_0, \ldots, \mathcal{M}_p$ using cross-validated prediction error, C_p (AIC), BIC, or adjusted R^2.

① GA (inspired Algos)

1. 种群
2. DNA




ATCG
(Base4)
0101 ......
Y
X1 ...... X300
表现型
[1 0011001]
基因型
1.2 53 3.7
MUTATION
(变异)
CROSSOVER
FITNESS
0.8
0.2
0.7
1.3
50
Population
[1001010 ..]
P=0.05%
step_size
f( )
R2

$\vec{X}_{30} \rightarrow \boxed{f} \rightarrow Y$ $X \; Y$

目的 $\vec{X}_{30} = ?$ Y 最大

$f(\vec{X})$

$1000 \left\{ [0.1 \;\; 1.2 \;\; \cdots \;\; 9.9]_{30} \right.$

N

$t \quad - \quad = \quad P_{t+1}$

R_{t+1}

$\Sigma \; 001001$

$\uparrow \qquad \uparrow$

$\sqsupset$

$\sqsubset \qquad P$

<u>GOAL</u>

x

max $\boxed{\text{FITNESS}}(x)$

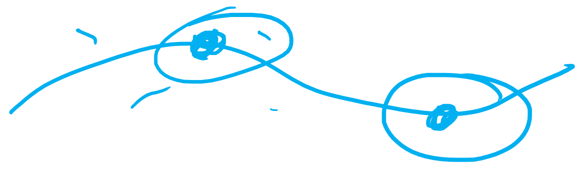
min x

$f(x^2)$

$100 - x$ $\boxed{10000}$

1	1
10	100
25	625
0 x^*	0

1. 概率.



	FITNESS
1 ___	1.2
2 ___	0.3
3 ___	0.1
4 ___	1.5
5 ___	
6 ___	

0 1

Random Num

$(r.n - 0.6) < 0$

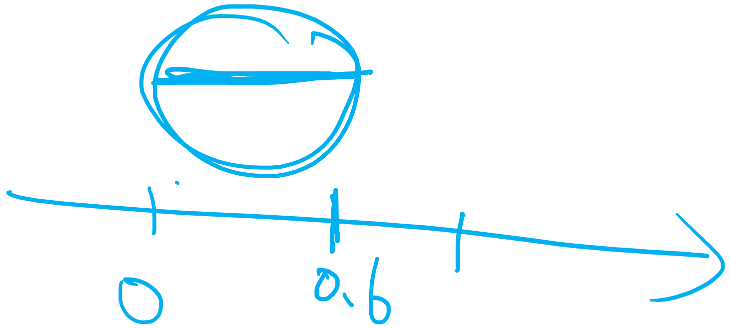
np.random.rand()

$\boxed{[0, 1]}$

P

$P_1 = \frac{f_1}{\sum f_i} = 0.6$

$P_2 = f_2 / \sum f_i$



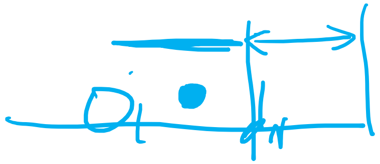
Implemetaion

Trick

60%

X_1

[ ]



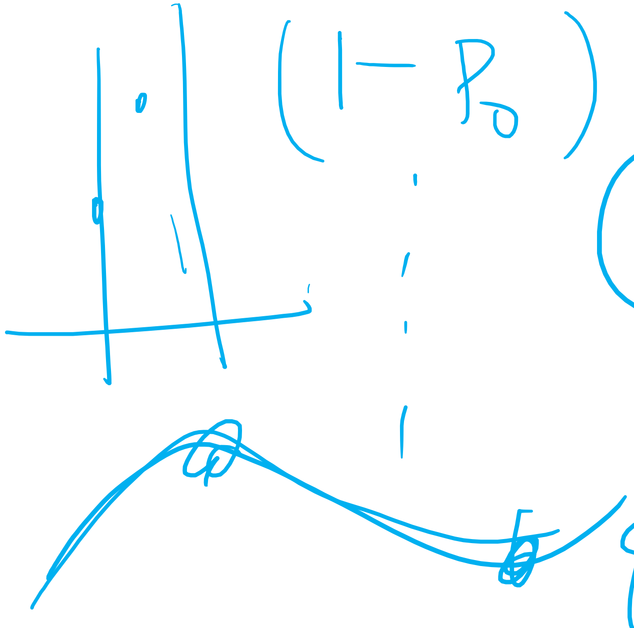
RANK.

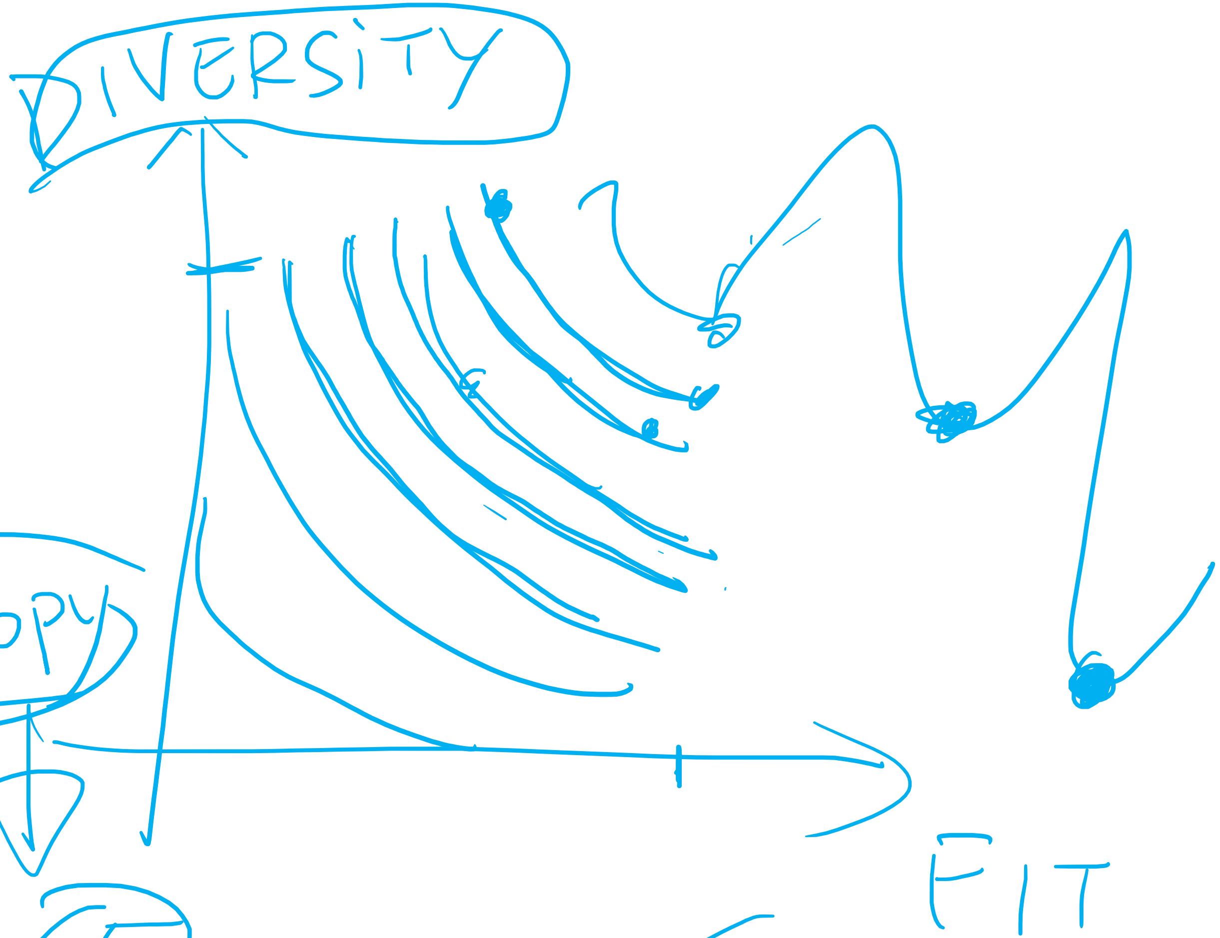
P_0

$(1 - P_0)$

var

Entropy





1 2 5

[ ]

"HELLO ␣ WORLD."

$f(s) = \#$ Char which is correct

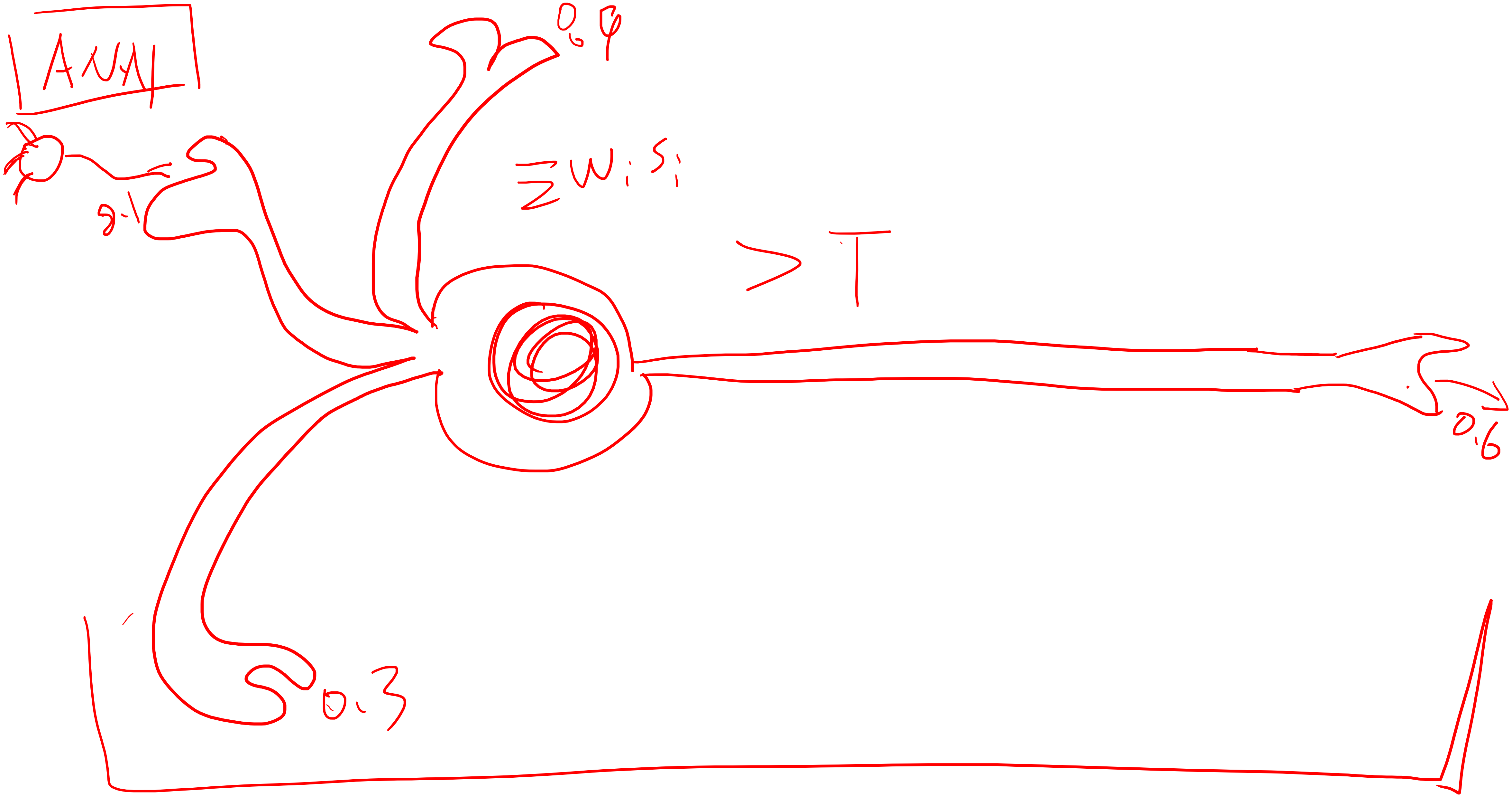
[$\boxed{01}$ $\boxed{25}$ $\boxed{16}$.

0 ~ 24

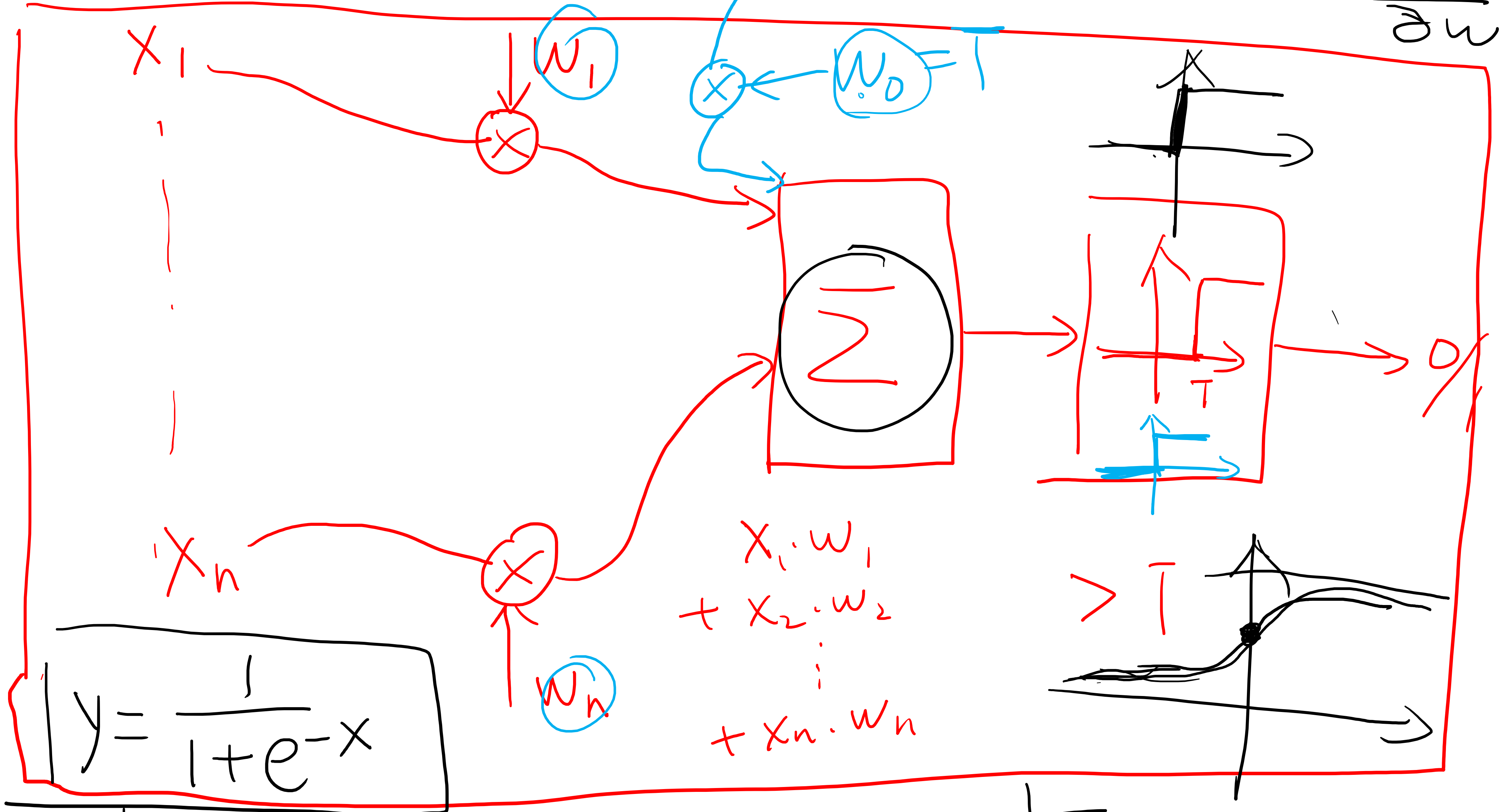
$y = x^2$

PPOS

NIPS


ANY
0.1
0.9
$\Sigma w_i s_i$
$> T$
0.3
0.6

X_{300} Y_1 bias term $\frac{\partial P}{\partial w}$

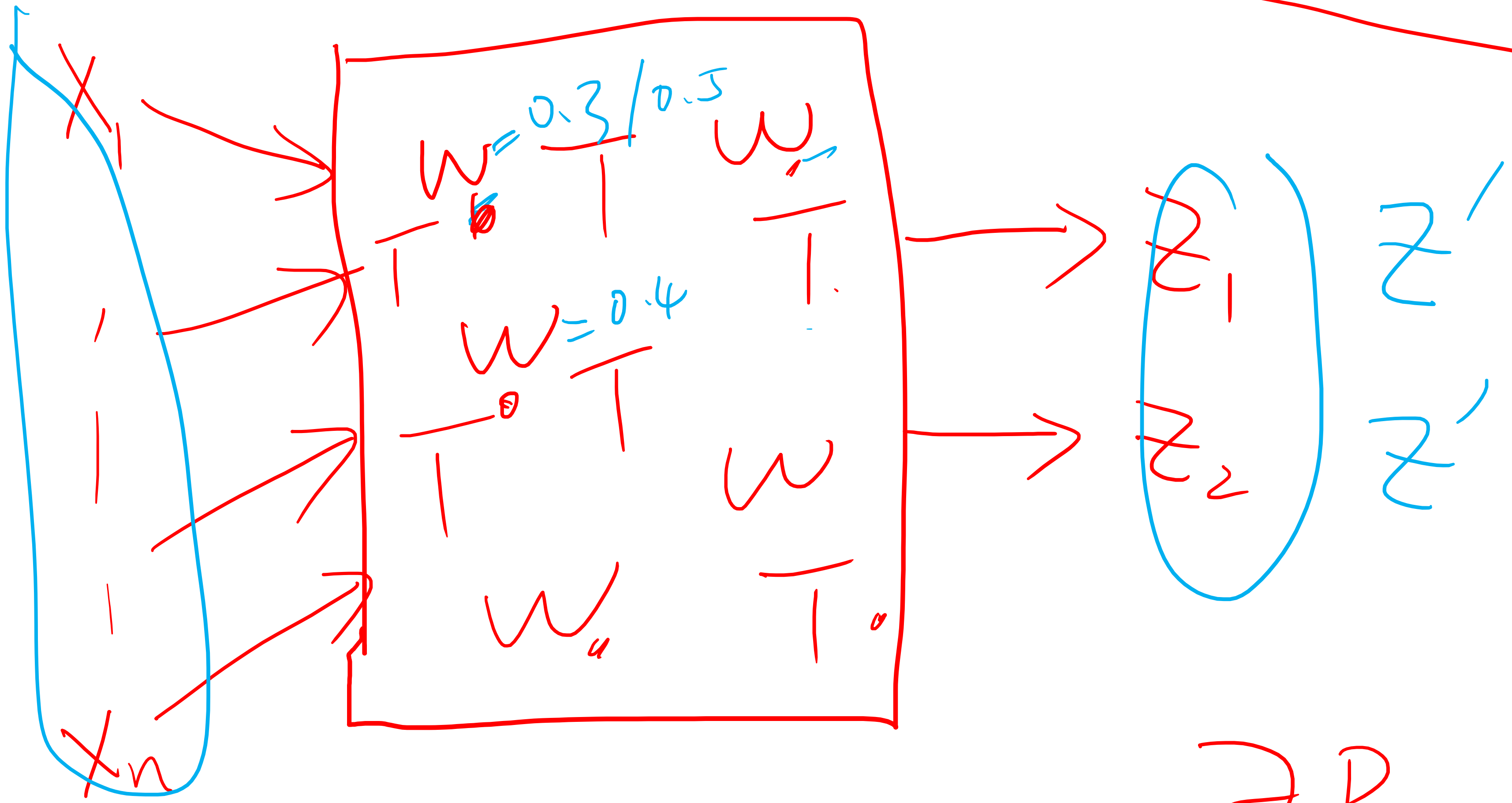


$$y = \frac{1}{1+e^{-x}}$$

$$y' = y(1-y)$$

$$\text{Sigmoid} = \frac{1}{1+e^{-\alpha}}$$

$$\boxed{\vec{z} = f(x, \vec{w}, T)}$$



$$\frac{\partial P}{\partial w}$$

$$\vec{z} = f(\vec{w}, \vec{x})$$

$$P = \|z - d\|$$

$$\boxed{\begin{matrix} w_1 = 0.2 \\ w_2 = 0.8 \end{matrix}}$$

$$\vec{d} = g(\vec{x})$$

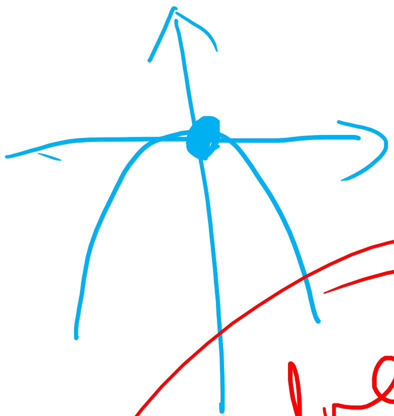
$$P = \frac{1}{2}\|\underbrace{\vec{z}}_{\text{真实}} - \underbrace{\vec{d}}_{\text{预测}}\|^2$$

$P = -\frac{1}{2}|z-d|^2$

w

$\frac{\partial P}{\partial w_1} = 1.2$

$\frac{\partial P}{\partial w}$



Andrew NG

w_1 0

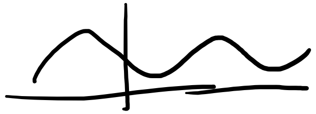
w_1 0

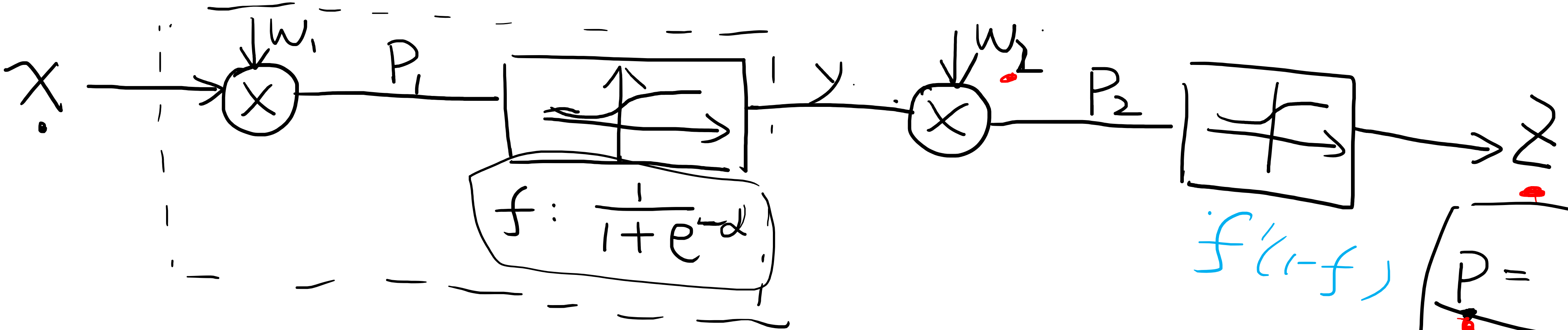
SGD

$$\Rightarrow \boxed{\overset{1.2}{w_1} \leftarrow \overset{1.2}{w_1} - \underset{0.03}{\overset{0.03}{\alpha}} \frac{\partial P}{\partial w_1}}$$

$$\boxed{X_{100\times 1} \qquad y_{1\times 1}}$$

$$f : X_1 \to y_1$$





$f'(1-f)$

$$P = (d-z)^2$$
$$= 2(d-z)$$

$$\frac{\partial P}{\partial w_2} = \frac{\partial P}{\partial z} \cdot \frac{\partial z}{\partial w_2} \rightarrow \frac{\partial z}{\partial P_2} \cdot \frac{\partial P_2}{\partial w_2}$$

$$\frac{\partial P}{\partial w_1} = \frac{\partial P}{\partial z} \cdot \frac{\partial z}{\partial P_2} \cdot \frac{\partial P_2}{\partial y} \cdot \frac{\partial y}{\partial P_1} \cdot \frac{\partial P_1}{\partial w_1}$$

$$P = (d-z)^2$$

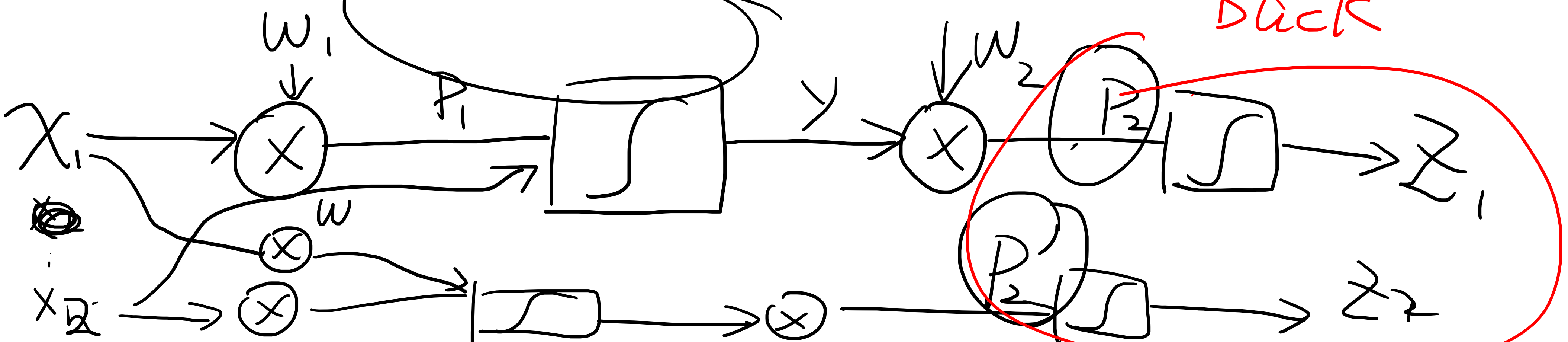
$$\frac{\partial P}{\partial z} = ?$$

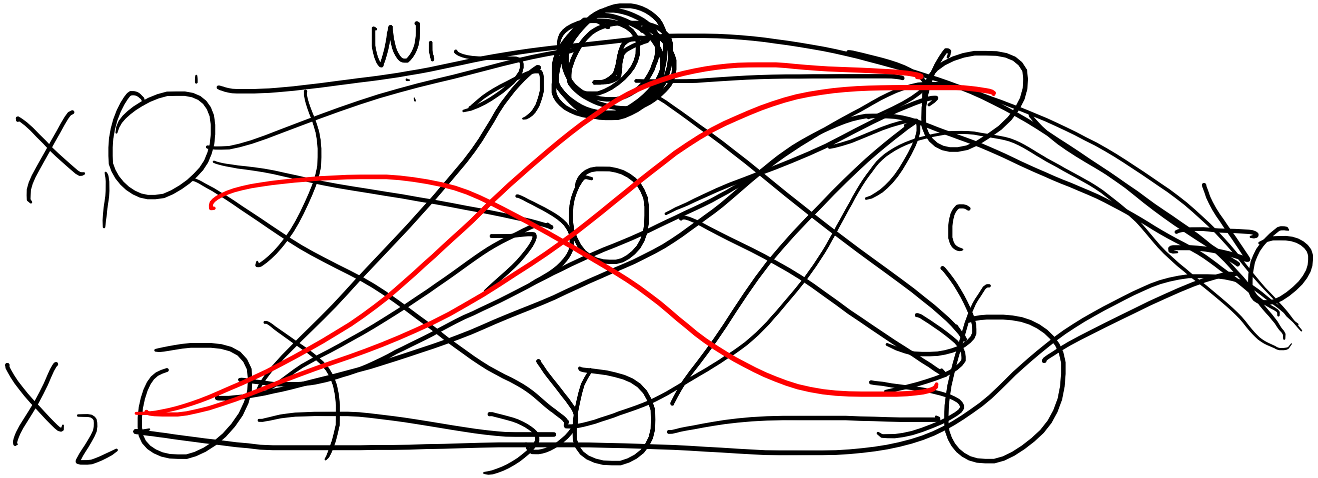
$$\frac{dP}{dz} = 2(d-z)$$

$$\frac{\partial P}{\partial w_2} = \frac{\partial P_2}{\partial w_2} \cdot \frac{\partial z}{\partial P_2} \cdot \frac{\partial P}{\partial z}$$

$$\frac{\partial P}{\partial w_1} = \frac{\partial P_1}{\partial w_1} \cdot \frac{\partial y}{\partial P_1} \cdot \frac{\partial P_2}{\partial y} \cdot \frac{\partial z}{\partial P_2} \cdot \frac{\partial P}{\partial z}$$

Back





$N \uparrow$ # Possible PATH $\uparrow\uparrow\uparrow$

O(# Hidden Layer)

S

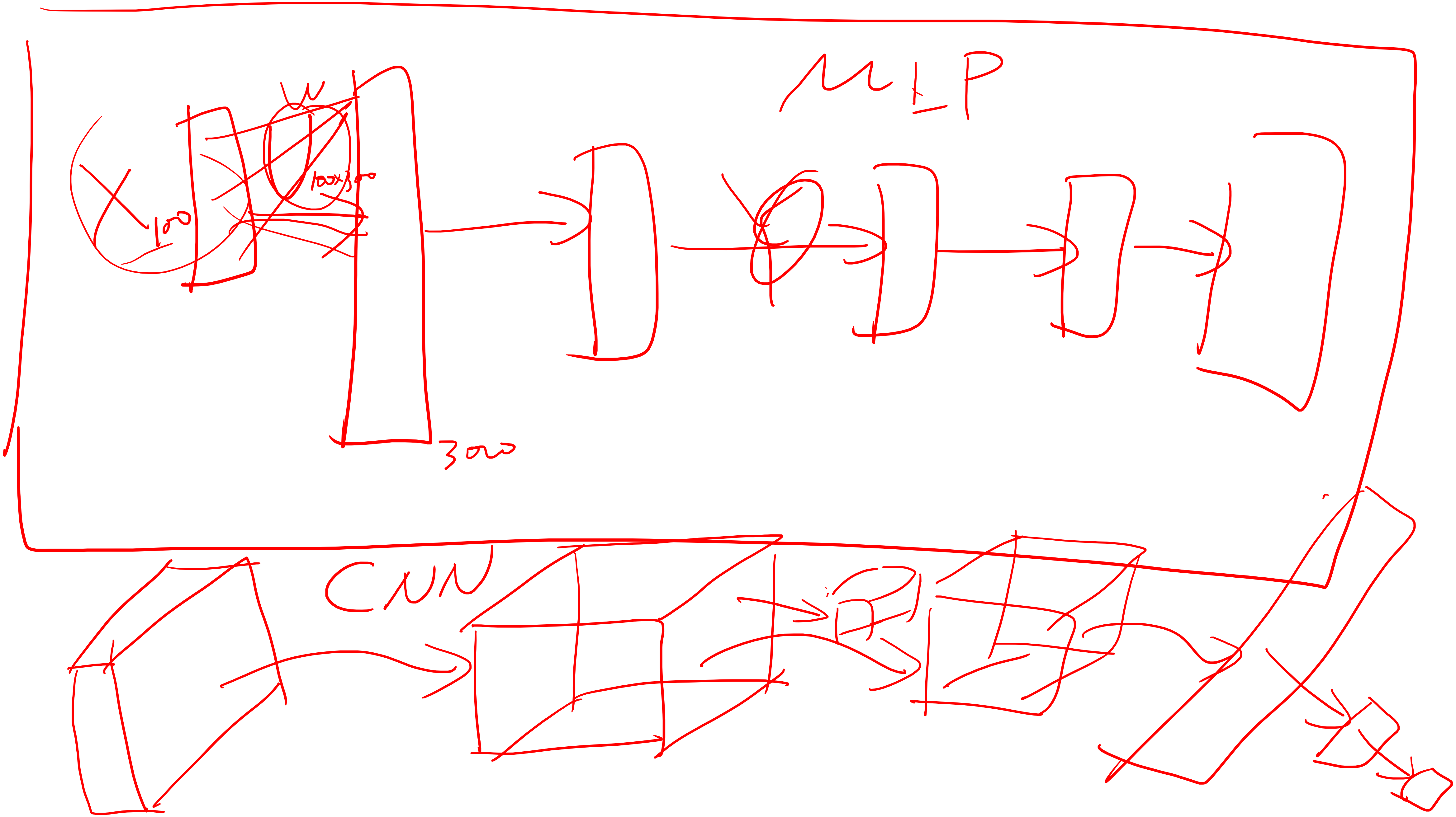
Re

递归

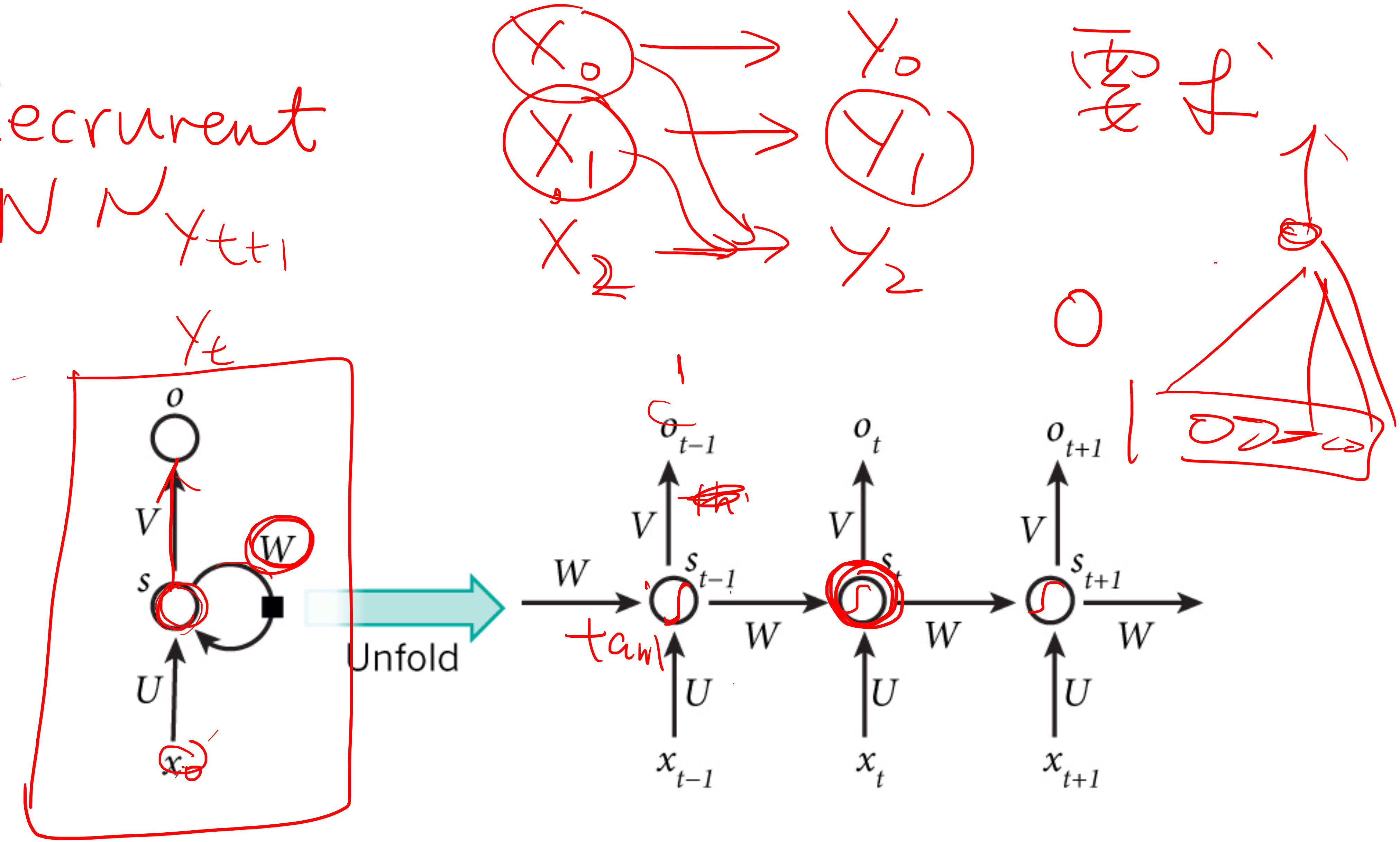
NLP → recursive NN

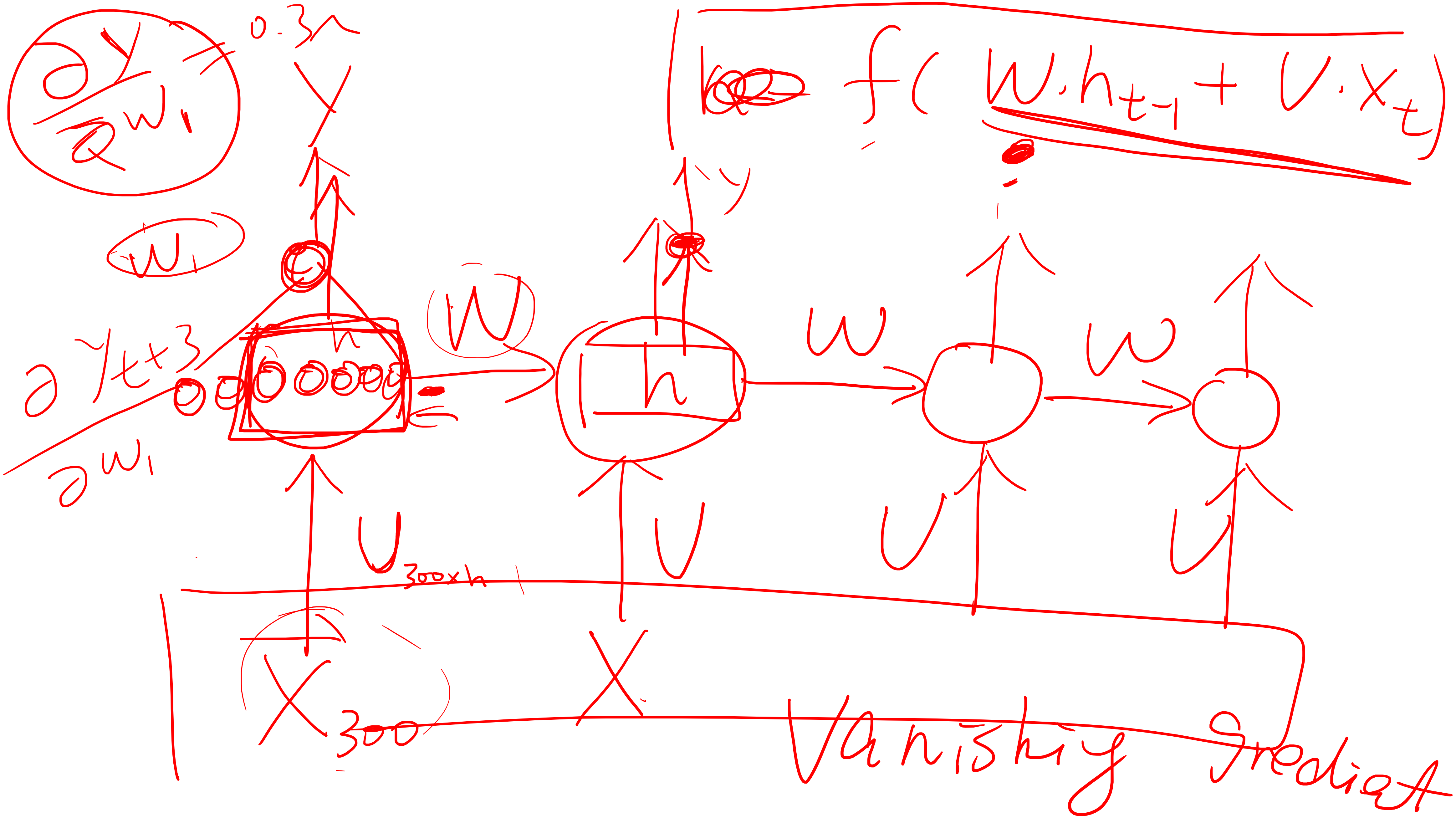
时间 → Recurrent NN

循环


MLP
X
100
W
100x300
300
CNN

Recrurent NN





$\frac{\partial y}{\partial w_1}$
0.3
y
w_1
W
w
w
h
h
$\frac{\partial y_{t+3}}{\partial w_1}$
U
$U_{300\times h}$
U
U
U
X_{300}
X
$f(W \cdot h_{t-1} + U \cdot x_t)$
Vanishing gradient

LSTM

4

Keras

LSTM(